Award Number: W81XWH-11-1-0744

TITLE: Development of Assays for Detecting Significant Prostate Cancer Based on Molecular Alterations Associated with Cancer in Non-Neoplastic Prostate Tissue

PRINCIPAL INVESTIGATOR: Farhad Kosari, Ph.D.

CONTRACTING ORGANIZATION: Mayo Clinic
Rochester, MN 55905-0002

REPORT DATE: Oct 2015

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*



| 1. REPORT DATE | 2. REPORT TYPE | 3. DATES COVERED |
|---|---|---|
| October 2015 | **Annual** | 30 Sep 2014 - 29 Sep 2015 |

**4. TITLE AND SUBTITLE**

Development of Assays for Detecting Significant Prostate Cancer Based on Molecular Alterations Associated with Cancer in Non-Neoplastic Prostate Tissue

**5a. CONTRACT NUMBER**
PC100553

**5b. GRANT NUMBER**
W81XWH-11-1-0744

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**

Farhad Kosari F, Cheville JC, Vasmatzis G, Karnes RJ, Manemann M, Murphy SJ

E-Mail: kosari.farhad@mayo.edu

**5d. PROJECT NUMBER**

**5e. TASK NUMBER**

**5f. WORK UNIT NUMBER**

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
**Mayo Clinic**
**Rochester, MN 55905-0002**

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

**10. SPONSOR/MONITOR'S ACRONYM(S)**

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

**12. DISTRIBUTION / AVAILABILITY STATEMENT**
Approved for Public Release; Distribution Unlimited

**13. SUPPLEMENTARY NOTES**

**14. ABSTRACT** The goal of this project is to develop biopsy based assays to assess the probability that patients with a negative biopsy or with a prostate cancer (CaP) Gleason score 6 (GS6) biopsy actually have "significant" CaP of Gleason score 7 or higher which was missed during the biopsy evaluations due to insufficient sampling. Experimental design includes discovery and validation steps. In the discovery step, we have generated a sizable set of close to 200 gene expression and methylation sequencing data from High grade PIN (HGPIN) collected by laser captured microdissection (LCM) and bulk normal samples. Our data set was further boosted through collaboration with another Mayo investigator. Through an RFA by the Mayo Clinomics Program, we completed gene expression and methylation sequencing of indolent and significant tumor samples (ca. 60 totals). Based on our previous experience, genomic profiles of tumor can provide crucial clues in "field effect" biomarker selection and therefore we anticipate these data will be important in the success of this project. We are very close to the completion of collecting RNA/DNA from the FFPE samples in the validation step. We expect that in the coming year we will be able to validate our findings in the discovery step and develop biopsy based models that will be useful in clinical settings.

**15. SUBJECT TERMS**
LCM AACR

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| | | | | | USAMRMC |
| a. REPORT | b. ABSTRACT | c. THIS PAGE | | | 19b. TELEPHONE NUMBER *(include area code)* |
| U | U | U | UU | 11 | |

**Table of Contents**

Introduction ----------4

Body / Results ---------- 4

Research Accomplishments ---------- 5

Additional related research activities ---------- 8

Reportable Outcomes---------- 10

Conclusions ---------- 10

References ---------- 11

**Introduction**:

Each year in the U.S. more than a million men with an elevated serum PSA or abnormal digital rectal exam undergo a prostate biopsy, and nearly 200,000 are found to have prostate cancer (CaP). Decisions to treat CaP are heavily influenced by the Gleason score (GS) of the tumor in the needle biopsy specimen. Gleason score is a measure of tumor differentiation based on the two most prevalent patterns of tumor growth. Patients whose entire tumor is composed of GS6 rarely progress, and recently, more men diagnosed with GS6 tumors on needle biopsy are selecting active surveillance rather than surgery or radiation therapy. In contrast, men with more poorly differentiated tumors (GS7 and higher) have a significantly increased risk of progression, and require treatment. Choosing the best treatment options for patients with biopsy GS6 is complicated by the fact that a biopsy procedure only samples a very small part of the prostate, and in about 30% of men, it underestimates the GS (1). In those cases, men with GS7 and higher ($GS7^+$) prostate cancer are assumed to have GS6 tumors potentially leading to inappropriate treatment. In addition, because of the limited sampling and 30% false negative rate for detecting cancer (2, 3), many men with a negative biopsy result may have clinically significant prostate cancer. Because of that, many of the 800,000 patients with a negative biopsy undergo repeat biopsies which can be frustrating for both patients and urologists. When a pathologist examines a prostate needle biopsy specimen, the focus is on the identification of prostate cancer and appropriate Gleason scoring. Very little attention is paid to the "normal" areas which often comprise the majority of biopsy samples. This is despite a considerable body of evidence suggesting that molecular alterations associated with tumor in adjacent non-neoplastic cells, the so called "tumor field effect", can provide valuable clues regarding the status of the tumor. Remarkably, the field effect alterations have also been associated with aggressive prostate cancer (4).

**Body / Results:**

The objective is to develop clinically relevant molecular models to predict significant prostate cancer with $GS7^+$ based on the prostate cancer field effect markers. This proposal will focus only on identification of significant tumors with $GS7^+$ because Gleason score is the single strongest predictor of outcome in men with prostate cancer, and has the greatest influence on the clinical management of men with prostate cancer. This proposal will concentrate on the "omics" areas where prostate cancer field effect has been best demonstrated, namely transcriptomic and epigenomics. There are two Aims. Aim I will identify and validate prostate cancer field effect markers associated with $GS7^+$ tumors. Aim II will develop and test molecular models for predicting upgrading in GS6 biopsies and for predicting $GS7^+$ cancer in a repeat biopsy.

**Table 1:** Bulk and LCM samples proposed in the original application (before 2014 modification) for the biomarker discovery step by the next generation sequencing

| Sample | Bulk | LCM (HGPIN) |
|---|---|---|
| **BP** | 5 | |
| **$N_{i6}$** | 5 | 5 |
| **$N_{3+4}$** | 5 | 5 |
| **$N_{8+}$** | 5 | 5 |
| **Total** | **20** | **15** |

Aim I will analyze 4 types of samples. These include non-cancerous tissues from CaP patients with (i) indolent GS6 CaP ($N_{i6}$), (ii) GS7 CaP ($N_7$), and (iii) GS 8 and higher CaP

($N_{8+}$). We also analyze benign prostate tissues from patients free of CaP (BP) as controls. Table 1 describes the sample sizes in the original application. BP samples are resected prostate tissues from patients who were not diagnosed with CaP but had their prostates resected in cystoprostatectomy operations because of bladder cancers. In the proposal for the first phase of the project, gene expression and epigenetic alterations are to be analyzed by next generation sequencing. Laser captured microdissection (LCM) is used to collect high grade prostatic intraepithelial neoplasia (HGPIN) lesions in 15 samples. The remaining samples are collected using bulk macro-dissection (Table 1).

**Research Accomplishments**:

Select NGS cases and collection of BP tissues: In April 2014, we requested and received approval for expanding the discovery set from the original plan. The new plan included 15-20 samples in each category. The revised plan allowed us to reduce the number of the FFPE samples in the first validation study to a manageable number for the size of this study. In coordination with the Tissue Request Acquisition Group (TRAG) at the Mayo Clinic, a process was implemented for the collection of resected prostates from cystoprostatectomy patients. Sufficient number of CPR samples for the discovery and validation studies were collected.

LCM collection (HGPIN) and bulk macro-dissection: Cases with adequately large areas of HGPIN for LCM and benign areas that did not include any preneoplastic or tumor regions by macro-dissection in each of the $N_{i6}$, $N_7$, and $N_{8+}$ categories were identified in our frozen prostate tissue registry. Under the supervision of Dr. Cheville, LCM samples have been collected by a meticulous and carefully implemented process to minimize degradation of nucleic acids, especially the RNA.

RNA and DNA isolation: Purified RNA from LCM samples by a Qiagen kit was of high quality (RIN numbers typically greater than 8) and quantity (tens of nano-grams). Also DNA from LCM samples by a Qiagen kit was of sufficient quantity for the RRBS library prep. DNA and RNA were simultaneously collected from the bulk samples by the AllPrep kit (Qiagen) and had very high quality and yield.

Sequencing library preparation: We have been working closely with the Mayo Genomics Facility (MGF) in preparing the libraries for RNA-seq. All LCM RNA-Seq libraries used 5-10 nanograms starting material and produced uniform output by the NuGen protocol. Similarly, all bulk samples used 100 ng total RNA and produced high quality TruSeq libraries. Furthermore, after testing various RRBS small sample protocols, we worked with the MGF in RRBS library preparation of LCM samples. All LCM and bulk samples produced high quality RRBS libraries.

Sequencing on Illumina HiSeq platform: Initial sequencing analyses used HiSeq 2000 platform. During the last year, MGF offered sequencing on the HiSeq 4000 platform with more than twice the throughput. We first ran pilot RNA-seq and RRBS testing on the new machine and with successful results continued with this platform in all subsequent experiments.

**Table 2:** Bulk and LCM samples collected in this study for transcriptome profiling. Numbers in parenthesis are data obtained through collaboration with Dr. Thibodeau. aN designates the epithelial benign samples collected by LCM.

| | Bulk | LCM | | |
|---|---|---|---|---|
| **Sample** | **benign Areas** | **HGPIN** | ***aN** | ***tumor** |
| **BP** | 19 (19) | | | |
| **$N_{i6}$** | 127 (120) | 18 | 6 | 20 |
| **$N_7$** | 87 (80) | 12 | 6 | 33 |
| **$N_{8+}$** | 12 | 18 | 5 | 5 |
| **Total** | 245 | 48 | 17 | 58 |

* Data generated through separate funding by the Mayo Center of Individualized Medicine and the Clinomics

Table 2 describes the samples with RNA-seq data. We collaborated with Dr. Thibodeau who also holds a DOD grant for transcriptome sequencing of bulk BP, $N_6$, and $N_7$ samples. This sharing of data greatly boosted our statistical power as Dr. Thibodeau's data included 120 $N_6$ and 80 $N_7$ cases. Table 3 describes samples with methylation sequencing data by RRBS. Both the RNA-seq (Table 2) and RRBS (Table 3) samples sizes are considerably larger than the initial study plans (Table 1). We anticipate that the expansion of the discovery set will greatly improve the selection of robust field effect biomarkers which will be validated in subsequent steps.

Analysis - Mapping of methylated DNA and transcriptome data: MGF pipeline for mapping of sequence data has been used. These pipelines were established through a rigorous bioinformatics analyses (5). Sequencing data for about 80% of the samples in tables 2 and 3 have been mapped. The remaining 20% are in the queue for sequencing and once sequence data are made available will be mapped using the same programs.

**Table 3:** Bulk and LCM samples with methylation sequencing (by RRBS) data

| **Sample** | **benign Areas (Bulk)** | **HGPIN (LCM)** | ***tumor (LCM)** |
|---|---|---|---|
| **BP** | 16 | | |
| **$N_{i6}$** | 20 | 19 | 20 |
| **$N_7$** | 5 | 7 | 20 |
| **$N_{8+}$** | 18 | 16 | 0 |
| **Total** | 59 | 42 | 40 |

* Data generated through separate funding by the Mayo Clinomics Program.

Selection of transcriptomic markers: We have been developing de novo bioinformatics algorithm that focuses in identification of recurrent skipped exons that distinguish $N_{8+}$ from $N_{i6}$. These biomarkers can be incorporated in sensitive PCR assays where even small fractions of cells expressing the splice variants can be readily detected. Table 4 describes list of candidates selected using this algorithm. Additionally, Table 5 describes additional candidates selected by the analyses of the transcriptome data that we have so far collected. These results will be updated once the remaining sequencing data are mapped.

**Table 4**: Selected candidates for testing by qRT-PCR based on de novo bioinformatics program developed in this research. These candidates provide targets for sensitive targeting skipped junctions.

| **Gene** | **Junction** | **BPC8+ (%)** | **CPR/BPC6 (%)** |
|---|---|---|---|
| **GBP3** | 5_7 | 67 | 22 |
| **MMP7** | 1_3 | 67 | 34 |
| **THBS1** | 9_11 | 92 | 32 |
| **ANKRD9** | 2_4 | 33 | 58 |
| **TNFRSF1A** | 4_6 | 75 | 41 |
| **EMILIN1** | 6_8 | 75 | 44 |

Case selection of the validation samples: The first validation step uses a similar sample size as in the discovery step. Twenty samples in each of CPR, $N_{i6}$, and $N_{8+}$ categories required for this step have been identified.

LCM Collection (HGPIN) and bulk macrodissection: Using an automated LCM machine, we have been collecting HGPIN and bulk areas of FFPE blocks from surgical cases. Of the 40 LCM samples needed, we have collected close to 30 samples and the remaining ones are scheduled to be processes in the next two weeks.

**Table 5**: Selected genes for further examinations Transcripts with lower (upper panel) and higher (lower panel) expression in $N_{8+}$Compared with $N_6$ were selected based on the analysis of the microarray data. Important cancer genes are shown in bold.

| **Lower in $N_{8+}$Compared with $N_6$** | | | | |
|---|---|---|---|---|
| AHA1 | CDH19 | CENTB5 | CLSTN2 | CYP4F8 |
| FABP4 | GOPC | GREM1 | HSD11B2 | **IDS** |
| KIAA1618 | LRRC31 | PCDH20 | PGC | PIP |
| TGM4 | **TP53** | WT1 | | |
| **Higher in $N_{8+}$Compared with $N_6$** | | | | |
| **BMP5** | DDX6 | F2RL1 | GDF15 | HAS3 |
| HFE | **JAK1** | **JUN** | KIAA1217 | **MET** |
| NAV1 | NPR3 | NR5A2 | OR51E2 | PGF |
| PPM1E | RND1 | SCGB1D2 | SLC2A3 | STC1 |
| TFF3 | TMEM37 | **TOP2A** | **WNT4** | |

RNA and DNA isolation and QC inspection: Using Qiagen AllPrep DNA/RNA FFPE kit, all collected samples have been processed and produced sufficient quantities of starting material for testing

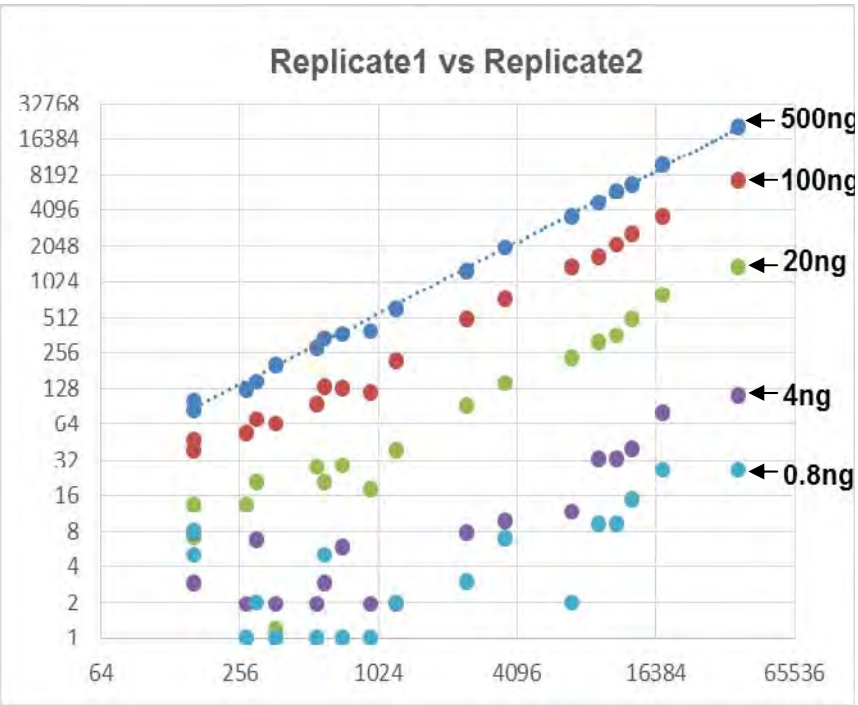
Quantitative PCR and methylation specific PCR: Several strategies for multiplex analyses of gene expression in FFPE samples were tested, including one-step and two-step RT-PCR, gene specific RT-PCR, and NanoString. At the end, the best performer was NanoString. With this technology, it was possible to produce results that were robust in less than 100 ng starting RNA (Figure 1). This sensitivity makes multiplexed gene-expression analyses in biopsies feasible as we typically obtained 100 ± 60 ng RNA (mean ± SD) by processing of the 2-3 sections of biopsy cores (data not shown).

Methylation specific PCR: A methylation specific PCR was developed and tested in FFPE samples for detection of GSTP1 methylation in prostate tumors (6). The assay was remarkably sensitive and capable of detecting methylation changes in tumor compared with normal samples in starting material compatible with prostate biopsies (5-10ng in our experience).

Biopsy Case selection (Aim 1.c. and Aim 2): Under IRB guidance we identified the prostate cancer biopsies that are needed for this study in the Mayo Clinic archives. Seven types of biopsy samples will be acquired.

1. Last biopsies from patients with multiple negative biopsies; Aims Ic (n = 50) and IIb (n=70). Our search identified 2308 patients who have had at least 2 biopsies at the Mayo Clinic.
2. Negative cores of Gleason score 6 (GS6) biopsies from patients identified with insignificant GS6 tumors after surgeries; Aim Ic (n = 100). Our search identified 5460 patients with a GS6 biopsy.

**Figure 1:** Nanostring experiments with varying starting RNA from FFPE blocks. Each point is the expression level of a gene (total of 18 genes). Blue, red, and green dots represent starting RNA of 500, 100, and 20 ng, respectively. Acceptable correlation was observed with starting RNA of tens (> 20ng) to hundreds of nano-grams




3. Negative cores of GS3+4 biopsies; Aim Ic (n = 50). Our search identified 1519 patients with GS3+4 biopsies.
4. Negative cores of GS8 and higher biopsies; Aim Ic (n = 50). Our search identified 763 patients with GS8 and higher biopsies.
5. Negative cores of Gleason score 6 (GS6) biopsies; Aim IIa (n = 700). Our search identified 5460 patients with a GS6 biopsy. Within all these specimens and data, from 2006-2007, a consecutive prostate needle biopsy and RP specimen Mayo database has been created for GS 6 tumors. This is comprised of 356 with an upgrading (GS>7) rate of 21% at RP. Upgrading occurs when the needle biopsy specimen is GS6 (and patient would be candidate for active surveillance) but RP specimen is GS7 or higher (indicating needle biopsy specimen sampling error). A separate more recent biopsy group is identified once again with paired biopsy and RP specimens at Mayo from 2010-2012 for validation work. This consists of 187 with an upgrading rate of 37%. Therefore 543 (356+187) patient biopsies have been characterized. We will characterize the remaining cases needed to reach 700.
6. Penultimate negative biopsies from patients identified with a GS6 cancer on the last biopsy; Aim IIb (n = 70). Our search identified 674 such patients.
7. Penultimate negative biopsies from patients identified with a GS7 and higher cancer on the last biopsy; Aim IIb (n = 70). Our search identified 232 patients. In this group, 72 patients would be 85 years or older by the second year of the study based on their age at their first visit to the Mayo Clinic.

**Additional related research activities:** In addition to the steps described above, we have engaged in other research activities which will enhance our abilities to accomplish PC100553 goals. These include:

Addition of a new member to the group: Prasuna Gupta, M.D. is a very diligent and excellent pathologist who is carefully going through the patients slides and selecting the right tissue blocks and biopsy cores for this project. Her contributions are invaluable to this research.

Genomic analyses of indolent and significant CaP: Based in our previous experience, field effect biomarkers can be robust if biomarker selection strategy is guided by changes in the corresponding tumor. In 2014, we secured a $50,000 award for generating transcriptome and epigenetic profiling of the iG6 and $G7^+$ tumors. This study will also help in developing a model to predict upgrading of GS6 biopsies and complement models based on FE markers in the benign tissue. As shown in Tables 2 and 3, we have close to 20 samples processed in each of significant and "indolent" tumor categories. Methylation and transcriptomic data from these samples are extremely valuable in guiding biomarker selection.

BPDE-DNA adducts as "field effect" biomarkers of "significant" prostate cancer: Previous studies by immunohistochemistry (IHC) have shown that the concentrations of BPDE-DNA adducts in the benign glands in prostate tissues that contain prostate cancer are correlated with the grade of prostate cancer ($p = 0.008$)(7). Prostates that contain high grade of CaP had higher concentration of BPDE-DNA adducts in their benign glands compared with prostates that had low grades of CaP. Therefore, these adducts can serve as "field effect" biomarkers of "significant" prostate cancer. We obtained funding through an NIH R21 mechanism to develop protocols for measuring these adducts in human tissues including prostates. We now have these protocols worked out and are planning to test them in prostate samples in the coming year.

Pending external applications: In a collaborative effort with engineers in the University of Illinois at Urbana-Champaign (UIUC), we have applied for an R33 which received a good score and was resubmitted for further consideration. This grant aims to develop models for stratification of indolent from significant CaP at the biopsy stage. In these applications we plan to develop sensitive nano-scale IR resonators to identify different cell types in a prostate tissue, including reactive stroma and myofibroblast which are modified in the CaP by the field effect. Additionally, selected biomarkers will be tested in a multiplexed fashion by quantum dots.

Collaboration with Dr. Thibodeau: Dr. Thibodeau at the Mayo Clinic also has a DOD award which requires transcriptome profiling of BP and $N_{i6}$ and $N_{4+3}$ and has graciously agreed to share his data in this project. As shown in Table 2, the additional samples through this collaboration will significantly enhance the number of samples for our analyses.

Un-anticipated events: In 2015, we spent efforts to test the new HiSeq4000 platform for RRBS and RNA-seq libraries. The new machine has significantly higher throughput than the previous generation (HiSeq2000) which allowed for multiplexing additional samples.

**Reportable Outcomes**

We have completed the most time consuming and challenging steps of this project which was the expanded sample collection for methylation and transcriptome sequencing. Additionally, we have prepared the DNA and RNA for almost the entire set of surgical validation cases and tested assays for the validation by gene expression and methylation experiments. Furthermore, through funding provided by the Clinomics program, we are in the final steps of sequencing “indolent” and “significant” tumor cases by RRBS and RNA-seq. Therefore, we are poised to harvest the results of a sizable set of data that we generated through the past years and we are confident that our selected candidate biomarkers will be incorporated in highly predictive models that can be applied to clinical settings.

We have reported on DNA rearrangements (such as translocations and deletions) in the CaP adjacent non-neoplastic cells in 2012 (8). Another publication (9) describes our previous work related to the CaP field effect. With the acknowledged support from this grant, this work was presented in the AACR meeting in April 2012. Finally, we have applied for a major grant initiative which if funded, will help expanding the scope of this study through the use of nano-scale devices and quantum dot based assays.

**Conclusions:**

We expect to incorporate the sizable dataset generated in this work in several publications. The first one which we plan to submit by the end of this year will assess the similarities of HGPIN and tumor genomic abnormalities in patients with indolent and significant CaP. Sequencing data is mostly available and the analyses pipeline established. Therefore these results are expected to be available for publication shortly. The other publication will be incorporating our *de novo* bioinformatics program for identification of skipped junctions in prostate cancer field. Remaining step for this publication will be validation of markers selected in Table 4. We expect this manuscript to be submitted by early 2016. Finally, by the middle of 2016, we expect to have an integrated gene expression and methylation based biomarker models tested in biopsies for distinguishing indolent and significant cancers. We expect these models will be the subject of significant publications and more importantly of use in clinical settings.

**References:**

1. Griffin CR, Yu X, Loeb S, Desireddi VN, Han M, Graif T, et al. Pathological features after radical prostatectomy in potential candidates for active monitoring. J Urol. 2007;178(3 Pt 1):860-3; discussion 3.
2. Patel AR, Jones JS, Rabets J, DeOreo G, Zippe CD. Parasagittal biopsies add minimal information in repeat saturation prostate biopsy. Urology. 2004;63(1):87-9.
3. Stewart CS, Leibovich BC, Weaver AL, Lieber MM. Prostate cancer diagnosis using a saturation needle biopsy technique after previous negative sextant biopsies. J Urol. 2001;166(1):86-91; discussion -2.

4. Malins DC, Gilman NK, Green VM, Wheeler TM, Barker EA, Vinson MA, et al. Metastatic cancer DNA phenotype identified in normal tissues surrounding metastasizing prostate carcinomas. Proc Natl Acad Sci U S A. 2004;101(31):11428-31.
5. Sun Z, Baheti S, Middha S, Kanwar R, Zhang Y, Li X, et al. SAAP-RRBS: streamlined analysis and annotation pipeline for reduced representation bisulfite sequencing. Bioinformatics. 2012;28(16):2180-1. PMCID: 3413387.
6. Baden J, Adams S, Astacio T, Jones J, Markiewicz J, Painter J, et al. Predicting prostate biopsy result in men with prostate specific antigen 2.0 to 10.0 ng/ml using an investigational prostate cancer methylation assay. J Urol. 2011;186(5):2101-6.
7. Rybicki BA, Rundle A, Savera AT, Sankey SS, Tang D. Polycyclic aromatic hydrocarbon-DNA adducts in prostate cancer. Cancer Res. 2004;64(24):8854-9.
8. Murphy SJ, Cheville JC, Zarei S, Johnson SH, Sikkink RA, Kosari F, et al. Mate pair sequencing of whole-genome-amplified DNA following laser capture microdissection of prostate cancer. DNA Res. 2012;19(5):395-406. PMCID: 3473372.
9. Kosari F, Cheville JC, Ida CM, Karnes RJ, Leontovich AA, Sebo TJ, et al. Shared gene expression alterations in prostate cancer and histologically benign prostate from patients with prostate cancer. Am J Pathol. 2012;181(1):34-42. PMCID: 3388167.